

Research Data Management in Policy and Practice: The DataRes Project

By: Spencer D. C. Keralis, Shannon Stark, [Martin Halbert](#), and William E. Moen

“Research Data Management in Policy and Practice: The DataRes Project,” (with by Spencer D. Keralis, Shannon Stark, and William E. Moen) in *Research Data Management: Principles, Practices, and Prospects*. Council on Library and Information Resources, 2013, pp. 16-38. URL: <http://www.clir.org/pubs/reports/pub160>



Research Data Management: Principles, Practices, and Prospects by the Council on Library and Information Resources is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#).

Abstract:

In this paper, we report findings of the DataRes Project, a two-year project funded by the Institute for Museum and Library Services (IMLS). We examine the perceptions of library professionals faced with supporting federal funding agency mandates for research data management plans, describe the state of data management requirements at major federal funding agencies, discuss our findings about the policy landscape at the top National Science Foundation (NSF) and National Institutes of Health (NIH) awardee institutions in the United States, and describe examples of robust responses to the needs of researchers for data management plan support.

Keywords: DataRes Project | research data management plans | data management

Article:

*****Note: Full text of article below**

Research Data Management

Principles, Practices, and Prospects

November 2013

OPEN ACCESS TO RESEARCH DATA IS CRITICAL FOR ADVANCING SCIENCE, SCHOLARSHIP, AND SOCIETY. RESEARCH DATA, WHEN REPURPOSED, HAS AN **ACCRETIVE VALUE**. PUBLICLY FUNDED RESEARCH SHOULD BE PUBLICLY **AVAILABLE** FOR PUBLIC GOOD. TRANSPARENCY IN RESEARCH IS ESSENTIAL TO SUSTAIN THE PUBLIC **TRUST**.

THE VALIDATION OF RESEARCH DATA BY THE PEER COMMUNITY IS AN ESSENTIAL **FUNCTION** OF THE RESPONSIBLE CONDUCT OF RESEARCH. MANAGING RESEARCH DATA IS THE RESPONSIBILITY OF A BROAD RANGE OF STAKEHOLDERS INCLUDING

The Denton Declaration:
An Open Data Manifesto



COUNCIL ON LIBRARY AND INFORMATION RESOURCES

Research Data Management

Principles, Practices, and Prospects

November 2013



COUNCIL ON LIBRARY AND
INFORMATION RESOURCES

ISBN 978-1-932326-47-5
CLIR Publication No. 160
Published by:

Council on Library and Information Resources
1707 L Street NW, Suite 650
Washington, DC 20036
Web site at <http://www.clir.org>

Copyright © 2013 by Council on Library and Information Resources. This work is made available under the terms of the Creative Commons Attribution-ShareAlike 3.0 license, <http://creativecommons.org/licenses/by-sa/3.0/>.



Cover art derived from Mariette Papić's poster for The Denton Declaration. Also included are images from Shutterstock.com.

Research Data Management in Policy and Practice: The DataRes Project

Spencer D. C. Keralis, Shannon Stark, Martin Halbert, and William E. Moen

Abstract

In this paper, we report findings of the DataRes Project, a two-year project funded by the Institute for Museum and Library Services (IMLS). We examine the perceptions of library professionals faced with supporting federal funding agency mandates for research data management plans, describe the state of data management requirements at major federal funding agencies, discuss our findings about the policy landscape at the top National Science Foundation (NSF) and National Institutes of Health (NIH) awardee institutions in the United States, and describe examples of robust responses to the needs of researchers for data management plan support.

Introduction

In October 2010, the National Science Foundation (NSF) announced its intention to require all grant applicants to include a plan for the retention and sharing of research data in their proposals, effective January 18, 2011. Such a plan—“a supplementary document of no more than two pages labeled ‘Data Management Plan’ ... [which] describe[s] how the proposal will conform to NSF policy on the dissemination and sharing of research results”—is to be included with every application for NSF funding, even if the plan is a statement that “no detailed plan is needed” (NSF 2013). Coming as it did amid the so-called Data Deluge, this data management plan requirement—often described by stakeholders as an unfunded mandate—initiated a furor across the academic world, from offices of research to research teams to academic libraries. Research universities across the United States are now struggling to develop consistent policies and programmatic implementations for institutional data management

functions. Research libraries and library and information science (LIS) programs in particular are scrambling to respond to these new requirements and to understand emerging requirements for curricula and training for both students and working information professionals. Recent surveys of the field and major white papers provide evidence that there is an acute need for research that will inform this process of curriculum and training development; research that documents the emerging patterns in data management policies; and research that documents the expectations of major stakeholders in the research cycle regarding data management roles, responsibilities, and professional training and preparation for those taking on data management responsibilities.

Funded by an Institute of Museum and Library Services (IMLS) Laura Bush 21st Century Librarians award, the DataRes Project was initiated at the University of North Texas to examine how research institutions responded to the NSF and other agency data management plan requirements in terms of policy and practical support for researchers, and to evaluate what role, if any, academic libraries and the LIS profession should have in supporting researchers' data management needs. The project, named DataRes as a shorthand mnemonic for the broad themes concerning research data that it examined, was a collaboration between the University of North Texas Libraries, the University of North Texas College of Information, and the Council on Library and Information Resources (CLIR).

Our research took place in a landscape that was changing as rapidly as things possibly can at the intersection of two monstrous bureaucracies—the grinding point where the tectonic plates of federal agency and academic administration meet. The most appropriate metaphor for the changes that we observed over the course of our research is neither the antediluvian hyperbole typical of discussions of “big data” nor the glacial or geologic metaphors usually applied to discussions of the academic and the federal bureaucracy. Rather, the changes we observed are evolutionary: slow, incremental change over time, punctuated by radical adaptations to local stimuli. Whether this evolution implies an aspect of survival of the fittest remains to be seen.

Background Survey

The DataRes Project developed in part as a response to a 2010 survey of library professionals at 200 U.S. research institutions. The survey, *Support for Research Data Management among U.S. Academic Institutions*, was an attempt to capture librarians' efforts and attitudes toward the management of research data and to determine the role of librarians in supporting data-intensive research in a digital environment (Moen and Halbert 2012).

To summarize the key findings of that survey as they are applicable to the present discussion, 100 percent of respondents (68 respondents, a response rate of 34 percent) believe librarians should “play a role in managing researchers' digital data.” Sorting the

possible roles that librarians may play in research data management into broad categories showed that a strong majority¹ of respondents believe that they should participate in the following aspects of managing data:

- informational (directing scholars to resources that will help them manage their own research data)
- instructional (providing training in the tools and information necessary for curating research data)
- infrastructural (providing space and resources for storing and accessing research data)
- cooperative (making tools and other resources available for scholars' use in managing research data)
- collaborative (actively participating in and guiding scholars' research data management)
- archival (preserving and providing access to research data once a scholar or research project no longer resides at the university)

A general concern of respondents, however, was the necessity of top-down institutional support, including financial support and adequate staff, to meet the needs of researchers in any of those roles. The following is a typical response:

While probably all of these [roles] are critical in terms of their usefulness to researchers, librarians would not be able to provide these services without substantive institutional support, so I have answered framed by the support for these services.

Other respondents cited “woeful budgetary times” to explain their libraries’ inability to provide data management support, although they acknowledged that such support is critical to the needs of researchers.

In terms of policy, respondents overwhelmingly responded in favor of an institution-wide research data management policy, with 78.2 percent of respondents describing such a policy as “very useful” (39.1 percent, 25 respondents) or “critical” (39.1 percent, 25 respondents). This finding led us to look closely at the policy landscape at top U.S. research institutions and to examine the ways in which libraries have responded so far to the data management needs of researchers.

Agency Guidance Documents

Our research began in July of 2011 with an environmental scan of the guidance for retaining and sharing research data at both the funding agency and institutional levels. We conducted focus groups at conferences and professional meetings with stakeholders in the

¹ The “strong majority” was nearly unanimous; the only responses of “negligible importance” were in the fields of cooperation (2 responses), collaboration (3 responses), and infrastructure (1 response).

research data management process.² We identified and compared the responses of academic libraries to the data management needs of researchers.

Among the federal funding agencies, only NSF, the National Institutes of Health (NIH), the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), and the National Endowment for the Humanities Office of Digital Humanities (NEH–ODH) have policies requiring plans for the retention and sharing of research data (Tufts University 2013). IMLS includes a questionnaire on the management of digital research products in its applications, but does not have a requirement for the retention and sharing of research data (IMLS n.d.).

One challenge in navigating the interagency landscape of data management requirements is that each agency maintains its own standards and formulas for grant applications. There is no consistency across agencies in data management guidance documents or even in general guidance for grant applications. As such, the documents we were able to examine from each representative agency varied, ranging from NSF's *Award and Administration Guide* (2011) to the Final NIH Statement on Sharing Research Data (2003). It is also difficult to find the authoritative document on a particular agency's policy. At NSF, each directorate and even individual program solicitations have specific requirements for the data management plan, and in a peculiar bit of circularity, the NSF policy on dissemination and sharing of research data referenced in the *Grant Proposal Guide* refers back to the *Grant Proposal Guide* "for full policy implementation" (NSF n.d., 2013).

As an heuristic exercise, we extracted the text of the data management plan guidance documents from NSF, NIH, and NEH–ODH and entered them into Wordles.³ The resulting word clouds tell a particular—and surprising—story about the priorities of each agency, and the thinking behind their policies.

The Wordle word clouds suggested that further analysis based on text mining could be fruitful. Text mining, or "distance reading," is a method of quantitative analysis of textual evidence, derived in part from the work of Franco Moretti and other scholars at the Stanford University Literary Lab (Moretti, 2011). Distance reading can make it possible to visualize patterns within texts or networks of associations among a corpus of texts that may be difficult or at least extremely time-consuming to see via close reading of individual texts.

² Focus groups were held on the following dates and locations:

- December 12, 2011, Washington, D.C. (at the Coalition for Networked Information winter meeting)
- January 20, 2012, Dallas, Texas (between the Association of Library and Information Science Educators and American Library Association midwinter conferences)
- June 27 and 28, 2013, Chicago, Illinois (during the American Library Association annual conference)

³ "Wordle is a toy for generating 'word clouds' from text that you provide. The clouds give greater prominence to words that appear more frequently in the source text." See <http://www.wordle.net/>. We limited the clouds to the top 100 words in each document and excluded commonly used words such as articles and prepositions.

National Science Foundation

In contrast, the NSF's *Award and Administration Guide*, Chapter VI.D.4 (figure 2) prioritizes “expected” and “Investigators,” but interestingly, the name of the agency is far and away the most prominent item in the word cloud (NSF 2011). This may indicate that the most important thing for the NSF was the mandate itself, not specifically the cultural implications (i.e., the benefit to the disciplines of such a mandate) or the practical implementation of the requirement.



Fig. 2. Wordle of the National Science Foundation's Award and Administration Guide. Chapter VI.D.4

With only 350 words, NSF's guidance to researchers is the smallest of the documents; it is, in fact, a small component of a larger document. However, it has the greatest vocabulary density (i.e., the greatest instance of unique words). “NSF” appears seven times in the document; “investigators,” five times; and “grantees,” “dissemination,” and “results,” four times each. There is no preponderance of usage of any of the key terms (“data,” “management,” or “sharing”) as in the other agency guidance documents. “Data” appears only three times. “NSF” occurs three times and is paired directly with “grants.” The focus, such as it is, appears to be on the authority of the granting agency. Interestingly, each directorate within the NSF gives supplemental guidance for applicants. Further analysis of these documents may be valuable for understanding the distinct ways in which these directorates are soliciting and evaluating data management plans.

At a focus group in December of 2011 with NSF program staff, the National Science Board, and research library administrators, NSF staff clearly articulated the importance of innovation in the disciplines' response to data management plan requirements. Although this approach accounts to some degree for the emphasis of the NSF policy, as well as the various directorate level instructions for plan development, researchers and library professionals at subsequent focus groups have offered other explanations, including a perceived unwillingness on the part of NSF or NSF peer review panels to make funding for data management support and repository services a part

of awards. This perception is contrary to the NSF's stated guidance to researchers that they should include costs for research data management in grant applications, and it derives largely from anecdotal information and library staff understanding of faculty priorities. The received wisdom among focus group participants was that faculty are simply unwilling to include these costs in their grant applications, and this filters up into peer review panels.

National Endowment for the Humanities

It is interesting to contrast NSF's policy with the word cloud generated from the executive summary of NEH-ODH's *Data Management Plans for NEH Office of Digital Humanities Proposals and Awards* (figure 3; NEH). In the NEH-ODH word cloud, "data" is extremely prominent, while "plan" is next in size, and "management" and "NEH" are roughly equivalent. For an agency serving disciplines that are largely perceived as not data-intensive, the focus on planning for the retention and sharing of research data is striking and indicates a shift in disciplinary priorities driven by the digital humanities.



Fig. 3. Wordle of the data management policy of the Office of Digital Humanities, National Endowment for the Humanities

The NEH document is the largest in the corpus, with 1,229 words, and has the lowest vocabulary density. "Data" appears 62 times in the document; in 9 instances, it occurs as part of the phrase "data management plan." "Management" appears an additional 11 times in the document (for a total of 20 uses), 8 of which are in the phrase "data management" on its own (as opposed to in the phrase "data management plan." The frequency with which the plan itself is mentioned—9 out of 20 uses—indicates a clear emphasis on the importance of the data management plan. Further, it emphasizes through repetition the research practice—data management—that the executive summary is introducing to the disciplines served by NEH.

Institutional Policies

In July 2011, the authors began an examination of published policies at the provost and office of research levels at the top NSF and NIH awardee schools. To develop the list of research universities for the institutional policy scan, we used the internal reporting tools from NSF⁵ and NIH. We selected these agencies because (1) our research agenda was a response to the NSF's requirement for research data management plans, and (2) NIH has the longest standing requirement for research data management plans. We set search parameters for the top dollar awardees for fiscal year (FY) 2010 and extracted the top 200 awardees from each agency. We synchronized the resulting reports, removing duplicates, stand-alone institutes, and individual awardees. The resulting list of 197 institutions constituted the pool for the policy scan.

We excluded IMLS because, while the agency does provide search capabilities for award information,⁶ it does not have an effective tool for extracting reports on awardees. Further, it does not offer explicit guidance on data management plans for applicants. IMLS requires applicants to complete a questionnaire, Specifications for Projects that Develop Digital Projects, about data practices, but does not require a data management plan per se on the model required by NSF, NIH, or NEH-ODH. Although we include the NEH-ODH guidance documents in our analysis for comparison purposes, because of the limited budget and scope of the Office of Digital Humanities, we have excluded those awardee schools from our scan (recognizing that many of those awardees may be captured in our report, regardless).



Figure 4: Count of institutions with data management policies

⁵ See <http://nsf.gov/awardsearch/>.

⁶ See <http://www.imls.gov/recipients/grantsearch.aspx>.

By performing Google searches using the institution names, “data management,” and “policy” as keywords, then duplicating the search using the institutions’ internal site search engines, we determined that only 18 percent (20 institutions) have publicly available policies requiring the retention and sharing of research data; the significant majority (82 percent) did not (figure 4). Many of the existing policies predate NSF’s requirement and were likely developed, at least in part, in response to NIH’s data management plan requirement. Institutions lacking a policy governing the retention and sharing of research data received in excess of \$13 billion in federal research funding from NSF and NIH in FY 2010–2011, a sizable investment of taxpayer money (Table 1).

Policy Found?	Sum of NSF	Sum of NIH	Sum of Total \$ Awarded
No	\$3,648,260,975.00	\$9,653,827,431.00	\$13,302,088,406.00
Yes	\$802,440,563.00	\$3,050,553,480.00	\$3,852,994,043.00
Grand Total	\$4,450,701,538.00	\$12,704,380,911.00	\$17,155,082,449.00

Table 1. Funds awarded to research institutions by the National Science Foundation (NSF) and the National Institutes of Health (NIH) in FY 2010–2011

University data retention policies tend to be fairly toothless, using statements of “recognition” of the importance of retaining and sharing research data or “encouragement” for researchers to share data rather than solid institutional mandates. One example is the policy of the University of New Hampshire (UNH), which states, “The University recognizes the importance of data sharing in the advancement of knowledge and education.” UNH goes on to restrict sharing of research data “only by specific agreement with persons or entities outside the University except where mandated by Federal funding agencies,” (UNH, 2012) further weakening the force of the policy. An index of known policies is available at <http://datamanagement.unt.edu/findings>.

Focus group respondents overwhelmingly supported the notion that if agency mandates are to be effectively implemented, institutional policy at funded universities will have to fall in line with agency priorities. There are myriad obstacles to this happening, but the greatest are institutional inertia and the liminal status of data as distinct research products. Focus group respondents from institutions with data management policies on the books reported that it can take as long as a decade to establish a provost-level policy. Respondents also described a state of affairs at many institutions in which offices of research are reluctant to engage with the policy or invest in the infrastructure necessary at least in part, for two reasons. First, they perceive the interest in data management as just a trend that agencies are not particularly serious about, and second, the return on investment is difficult to calculate. Offices of research would rather wait until agencies issue a more solid mandate than invest in data services and infrastructure now.

Further, the existing institutional policies are weak, and compliance tends to be limited because the only way to compel faculty to adhere to such a policy is to make compliance a mandate for tenure and promotion, a step no institution is willing to take. Focus group respondents uniformly reported that researchers tend to be reluctant to share data, considering them either residual products of their research or something so idiosyncratic, specialized, or proprietary that they simply prefer not to share the data. Further, data as such are neither valued nor rewarded as research products for tenure and promotion, so they will not be a priority for research faculty whose efforts are focused on publication and the next grant application.

Of those institutions lacking publicly available policies for data management, it is possible that some have such policies, but that they are not public-facing. It is also possible that some institutions are in the process of revising their data management policies or drafting new policies in response to the demands of NSF and other funding agencies. However, given the pace of change at most institutions, it may be years before new policies are implemented.

Data Management in Libraries

Unsurprisingly, given the emphasis of federal agencies on the data management plan itself, many efforts at both the library and institutional levels have focused on support for researchers writing their plans rather than on implementing the plans. For example, as of this writing, more than 100 institutions are registered with the DMPTool, meaning that they have Shibboleth login access to the tools for local researchers to develop plans, as opposed to the eight contributing institutions working on development of the tool (California Digital Library, 2013a). Although this focus is certainly important and reflects the short-term needs of researchers, it does not address what is necessary to implement a data management plan. Development of resources for long-term preservation and access to research data has been uneven and is generally less robust than support services for plan development.

In the course of our research, we identified 32 universities where libraries are providing some level of data management plan support for researchers, but this number is far from comprehensive. Models of support vary widely, from simple web pages linking back to the policy and guidance documents of federal funding agencies, to programs that offer workshops and other practical support for researchers, to infrastructure projects costing millions of dollars per year, or a combination of these. At the University of Minnesota, for example, the libraries provide a range of data management support functions (University of Minnesota 2011). Library specialists can help draft data management plans, consult on funding agency requirements, confer on subject-specific data repositories, and give access to on-campus research computing resources. In collaboration with the university office of research, the library also offers data management workshops to graduate students, faculty, and researchers. These

workshops provide continuing education credit, a requirement for principal investigators. As of August 2010, 250 faculty members had participated in the workshops and consultations, and six departments had invited librarians to conduct workshops for their entire staff (Kelley 2011).

To highlight the diversity of responses to the demand for data management services, the DataRes team organized a panel titled Meeting the Challenge of Data-Management Support in Academic Libraries, for the EDUCAUSE conference in November 2012 in Denver, Colorado. The panel featured Michael Witt, assistant professor of library science at Purdue University; Deb Morley, head of specialized content and services at the Massachusetts Institute of Technology (MIT) Libraries; Sarah C. Williams, life sciences data services librarian at the University of Illinois at Urbana-Champaign; and Ardys Kozbial, chair of the Data Curation Working Group at the University of California–San Diego Libraries. Panelists discussed their libraries' interventions in data management support for researchers at their institutions. The panelists' presentations reflected the findings of the DataRes study, indicating diverse responses to data management ranging from robust, infrastructure-driven models to ad hoc support provided by individual librarians, depending on the resources and culture of a given institution. The panel was broadcast live from Denver as part of the EDUCAUSE online conference, potentially reaching an audience of thousands of participants.⁷

In the libraries at the University of Illinois at Urbana-Champaign, the liaison librarians serve the disciplines most affected by the NSF mandate, and they drive support for data management. The life sciences librarian developed a links web page to give researchers access to information from funding agencies, information about data repositories, and a list of services, including help in developing a data duration profile for research projects using Purdue's Data Curation Profile Toolkit (originally developed in partnership with the UIUC Graduate School of Library and Information Science) (Purdue University n.d.). The library support page includes a link to the Illinois Digital Environment for Access to Learning and Scholarship (IDEALS), the institutional repository, which treats data sets as digital objects, but is not explicitly a data repository (University of Illinois at Urbana-Champaign n.d.).

The MIT Library publishes self-help resources for researchers, including checklists for data management plans, and advice on metadata, file formats, and data security. Specialist librarians also provide consulting services for researchers to help with the development of data management plans and the preparation of data for subject-based and institutional repositories (MIT n.d.b). The data management resources page links to DSpace@MIT, the university's institutional open access repository, which is a service of the libraries. DSpace@MIT is described as "stable, long-term storage for their

⁷ Details on the panel, including presentations, may be found at <http://www.educause.edu/annual-conference/2012/meeting-challenge-data-management-support-academic-libraries>.

digital research and teaching output and to maximize exposure of their content to a world audience” (MIT n.d.a). Although it is not exclusively a data repository, DSpace@MIT does support data sets.

The library at the University of California at San Diego offers a suite of data management services under the umbrella of Research Data Curation Services (University of California at San Diego 2013b). Data curation can best be understood as a life cycle approach to research data management that includes planning, the research process itself, preservation and access, and reuse or deaccession. Assistance is provided for the development of data management plans through individual consultations. The library supports long-term preservation in collaboration with the university’s Research Cyberinfrastructure (University of California at San Diego 2013a) and the University of California Curation Center, part of the California Digital Library (2013c). The EZID service of the California Digital Library supports digital object identifiers for data and other digital content, thus allowing researchers to create identifiers and to assign and store citation metadata for digital objects (California Digital Library 2013b).

Purdue University offers a range of services centered on the Purdue University Research Repository (PURR). A platform for life cycle data management, PURR provides data management plan development with boilerplate language, collaboration space for projects, digital object identifier service for data sets, and long-term preservation and access to data sets (Purdue University 2013). The suite of services is free to Purdue faculty and graduate students, with nominal costs for projects requiring storage above a standard set of space thresholds. Technologically advanced and infrastructure-intensive, PURR is an exemplary suite of services; however, an institution lacking Purdue’s financial resources would find it impossible to replicate.

As we shall see in the analysis of our two surveys, most of the funding for research data management support is coming from libraries themselves, with little or no financial support from offices of research, indirect funds, or other university sources. At one focus group, a visibly irritated program officer declared, “I don’t know what you all [librarians] are complaining about. We’re sending you business.” But this “business” is not the sort that pays. Faculty are accustomed to free library services; in a time when library resources are diminishing, any additional services to meet the needs of research data management requires cuts in traditional library services, such as subscriptions, book purchases, and student services.

Even at the institutions that have been noted, the commitment of the libraries to research data management varies widely. In most cases, either intra-institutional or extra-institutional collaboration has been key to providing research data management support. The level of support that libraries offer is largely contingent on the commitment of both the library and the university administration to provide financial and staff resources. In some instances, a simple links web page is the only intervention sanctioned by university or library administrations unwilling or unable to invest in the infrastructure necessary for long-term preservation and access to research data.

Primary Survey

To identify the current trends in research data management at research institutions, we distributed an online questionnaire, titled the DataRes Online Survey (DROS), in an early stage of the study. The data collected supplies substantial evidence to support the previous findings from the policy scan and corroborates testimonials from the focus groups. Participant responses also influenced the development of a secondary survey that was distributed a year later. Those results are reported in the next section.

The policy scan had indicated a significant lack of institution-wide policies in the top awardee research institutions, and this finding was loosely supported by the DROS (n=231), in which we asked various stakeholders if their institution had a policy governing the retention and sharing of research data (figure 5). To clarify the term *stakeholders*, participants defined themselves as librarians, research faculty, archivists, data managers, deans, and students, among several other professional titles. The survey instrument can be found at <http://datamanagement.unt.edu/findings>.

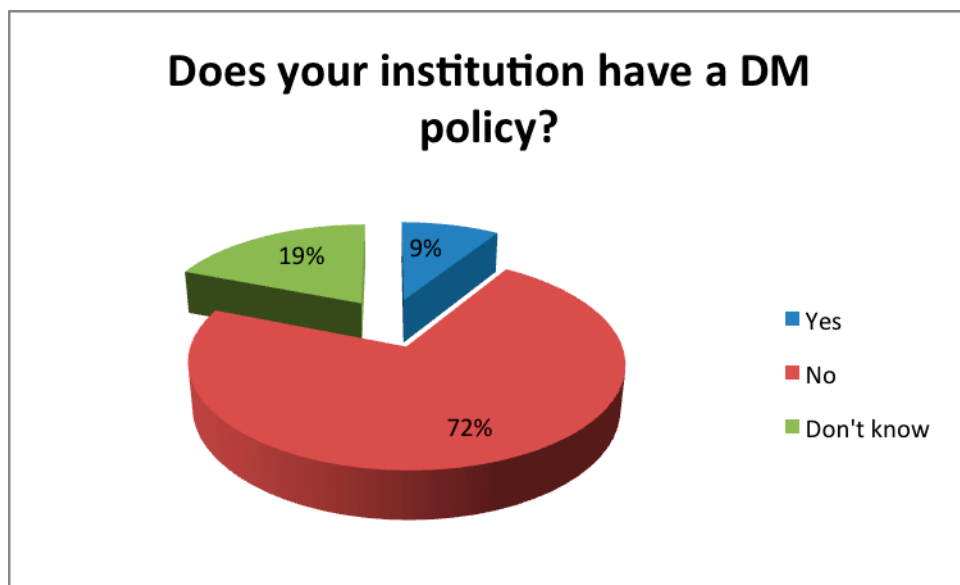


Fig. 5. Participant response to the question, "Does your institution have a policy governing the retention and sharing of research data?"

Only 9 percent of participants answered "yes," while the majority (72 percent) reportedly were employed by or enrolled at an institution that does not currently have a policy. An alarming 19 percent claimed, "I don't know," which could be equated with a "no" response, because the participants' lack of knowledge could suggest that even if a policy were in place, it is not being enforced to a degree that requires awareness or procedural changes.

We applied no mechanism in the survey to prevent multiple individuals from single institutions to respond. Consequently, we expect the percentages for this particular query to reflect higher percentages than those that actually exist. We felt that because most of

the questions pertained to individual preferences and experiences, such a limitation would have hindered our results more than it would have helped overall.

Immediately following this question on the presence or absence of a policy, we asked the participants to indicate how strongly they agreed or disagreed with the following statement: “I believe that an institution-wide data management policy is valuable” (figure 6).

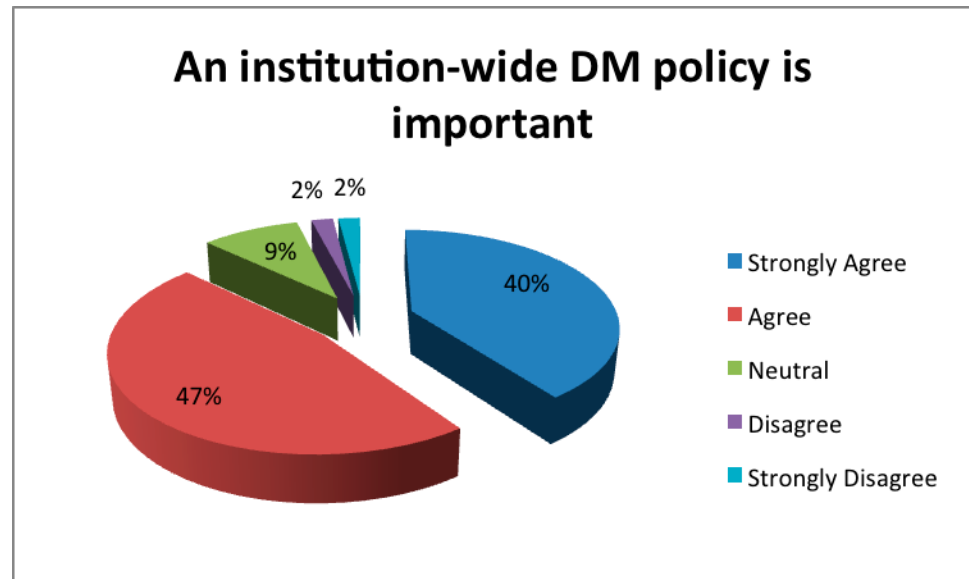


Fig. 6. Participant response to the statement, “I believe that an institution-wide data management policy is valuable.”

The majority (87 percent) indicated either agreement or strong agreement with the statement, while only a combined 4 percent disagreed or strongly disagreed. The remaining percentage showed a neutral opinion on the subject. These responses suggest that stakeholders are eager to see their institutions make a clear proclamation on the subject of research data management, which is consistent with the responses that we have received from focus group participants. It also invites a more complex conversation on policy enforcement, support, and the infrastructure required for retention and sharing of data. In this initial survey, we focused our inquiries on support and infrastructure to establish a baseline understanding of how institutions currently handle these needs.

As a starting place, we asked participants where their data are physically located, and more than half the respondents reported that data were kept on a “local computer or external hard drive” (54 percent). On a follow-up question, 72 percent said that they would use institutional repository services if they were offered. Table 2 gives more detailed information on the desired data management services, breaking down specific needs and indicating the departments that the respondents believe should be responsible for providing aid.

With the exception of “data storage infrastructure,” which was viewed as a responsibility of the information technology services

department, the majority of the participants indicated a preference for repository services to be provided by the office of research or the library. Table 2 is particularly interesting because of the implications for collaboration among departments. The spread of responses suggests that researchers view different aspects of data management as falling under different offices' expertise and that a collaborative approach across multiple departments and offices may be the best way to provide the most desired services.

#	Question	Office of Research	Library	Your Department	Schools of Library and Information Science	School of Computing	Information Technology Services Department	External Service Provider	Responses
1	Workshops on best practices for data management	58	85	13	7	5	26	12	206
2	Workshops on preparing data management plans for funding agencies	74	70	12	7	2	9	7	181
3	Templates for data management plans for funding agencies	73	68	12	5	2	15	9	184
4	Assistance composing data management plans for grant proposals	75	71	14	5	3	7	7	182
5	Data storage infrastructure	22	63	13	5	6	79	15	203
6	Other data management services provided by your institution	18	34	9	4	2	17	9	93

Table 2. Participant response to the question, "If your institution offered the following services and resources, would you take advantage of them? If so, please indicate in which department or office you believe these services and resources should be based. If you don't believe you would use the service, please leave that row blank."

Secondary Survey

To delve deeper and address gaps identified in the first survey, we developed a secondary survey to be sent to vice presidents of research, deans, and higher-level administrators. Although we felt that the first survey had more than addressed the perspective of librarians, we were dissatisfied with the response rate from individuals in administrative positions. Because people in these positions would drive any policy change, we felt it necessary to target them specifically with the DataRes Administrator Online Survey (DRAOS).

We also hoped to gain a better picture of what changes, if any, we could expect to see in the future through the administrators' reports of current planning and priorities. To administer the survey, we assembled a list of 400 contacts at the institutions from the prior sample and e-mailed them directly with the survey link. For 45 days, the link was left active, and responses were accepted. At the end of that period, 33 complete survey responses were collected. Figure 7 shows the makeup of our sample, according to the way in which the individual respondents described themselves.

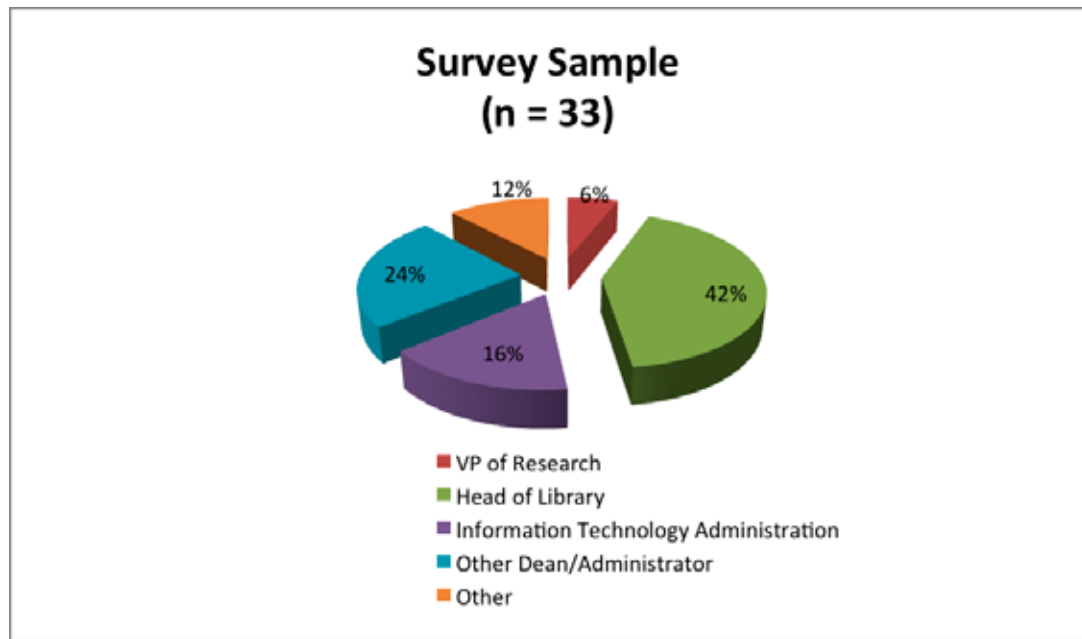


Fig. 7. Administrator responses to the question, "Which of the following best describes you? You may choose more than one."

University librarians, deans of libraries, and library directors made up the majority of our respondents, which we grouped together in the more general category of "Head of Library." The second largest group, "Other Dean/Administrator," included assistant deans and directors from university libraries and graduate schools. Those who defined themselves as "other" were primarily librarians; this group made up only 12 percent of the group, and these titles were often secondary as we allowed individuals to select more than one descriptor, should it apply.

We repeated in the secondary survey several of the questions from our initial survey to compare the librarian versus administrative perspectives. For example, we asked in the first survey which services and resources, if offered, the respondent would take advantage of, and most of our answer choices revolved around the early stages of data management and plan development. When addressing the administrators, we asked them, "At which stage of the data management process do you feel the average researcher at your institution needs the most support?" Figure 8 illustrates the administrators' responses.

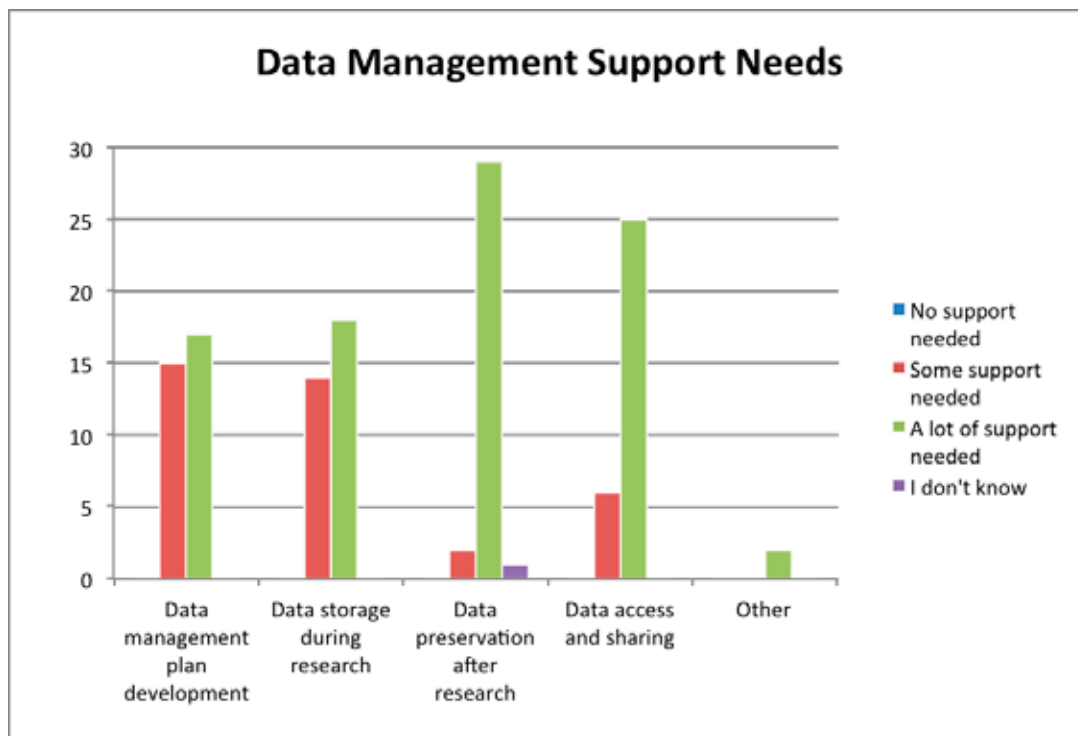


Fig. 8. Administrator responses to the question, "At which stage of the data management process do you feel the average researcher at your institution needs the most support?"

In general, administrators indicated that their researchers needed support at all stages, but they emphasized support during the later stages of "data preservation after research" and "data access and sharing." The two written responses for "Other" for which "a lot of support" is needed were "personal archiving" and "data citation." No responses indicated "No support needed" for any service. Recalling our responses to the DROS, slightly greater numbers indicated a preference for help in the form of "workshops on best practices for data management" and "data storage infrastructure." Meanwhile, fewer people indicated an interest in data plan creation and templates. Not only are the results of the DROS and DRAOS consistent with each other, showing that administrators and librarians at least have similar perspectives, but they also contradict some current practices in the field, which place high emphasis on helping researchers at the earliest stages with data management plans.

In exploring this contradiction, it is important to keep in mind that helping with data management plans and providing workshops for plan development are possibly the easiest and least costly response to the new expectations for data management from the federal granting agencies. Preservation, access, and sharing take additional infrastructure, expertise, and quite a bit more effort and cooperation between researchers and service providers. Answering the questions of how to store data long-term, where to put it, and how long to keep it are far more difficult than organizing hour-long workshops on what a data management plan is. Furthermore, because they were the low-hanging fruit, more early-stage services have

already been created and offered at research institutions, so now it is the later stages that are matters of concern. Whatever the explanation, both of our surveys suggest that late stage data management infrastructure and education are in high demand.

To determine a realistic baseline for the current responses of research institutions to data management needs, we asked in the DRAOS for respondents to report on what services their institutions currently provided and where those services were primarily based (table 3).

#	Question	Office of Research	Library	Individual Departments	Schools of Library and Information Science	School of Computing	Information Technology Services Department	External Service Provider	Total Responses
1	Workshops on best practices for data management	9	19	6	2	0	3	1	40
2	Data storage infrastructure	4	17	15	1	2	25	6	70
3	Continuing education courses on managing research data	1	3	3	4	0	1	0	12
4	Institutional repository for research	3	25	1	0	0	6	1	36
5	Administrative staffing to provide guidance and information regarding research data management	15	23	6	0	0	10	0	54
6	Other	6	8	2	0	0	4	0	20

Table 3. Administrator responses to the question, "What services, if any, does your institution currently offer to manage the retention and sharing of research data? Please indicate all services offered at your institution, and which departments oversee those programs. If a service is provided by more than one department, please indicate all departments involved."

The results showed an encouraging amount of overlap. The most offered service was data storage infrastructure, which was the second most desired service, with only a very small gap between it and the first most desired service (3 votes; see table 3). Also, the administrators reported that such infrastructure is housed primarily in the information technology services department, which was the DROS respondents' preference. For the most desired service, "workshops on best practices for data management," respondents indicated that libraries are the largest single providers of that service, but there were fewer overall offerings available for these workshops than other desirable services. (Note: DRAOS also does not account for multiple individuals responding from single institutions.) This is

a noticeable gap that could be a starting place for institutions hoping to begin providing aid to researchers in this area or for those already providing some services and looking for opportunities to do more.

The spread of services across various departments also echoes the DROS. The library is portrayed as a primary setting for all services except “data storage infrastructure” and “continuing education courses on managing research data,” but a notable number of respondents reported many services housed in the office of research and other departments as well. The education courses show up quite evenly distributed between the library, individual departments, and schools of library and information sciences, while the storage infrastructure is housed mostly by the information technology department. Again, this demonstrates an opportunity and need for collaboration across campus, which is currently being fulfilled at some of these institutions.

Next, we wanted to explore how administrators were handling the financial aspect of these new mandates. In the DROS, we had asked, “Do you typically allocate financial resources in your grant proposal budgets for data management?” Only 24 of the DROS respondents (13.95 percent) said “yes” to this question. Curious for more detailed information, we asked the administrators, “Does your institution allocate financial resources for data management?” Then we immediately proposed the follow-up, “From which sources are these funds drawn?” The results appear in figures 9 and 10, respectively.

Only half of the respondents were from institutions that offered financial resources for data management, and of that half, only 17 percent reported that principal investigators included data management in their grant proposal budget. Still, this number was higher than the 13.95 percent who said they did so in the initial survey. The majority of respondents (31 percent) indicated a hybrid model of funding, drawing from a mixture of all sources. Also important to note is that twice as much funding apparently comes from the library budget than from any other department budget.

Discussions in focus groups reflected this state of affairs; participants reported that vice presidents of research and other administration officials are reluctant to commit funding or other institutional resources to research data management support. Participants described conversations with administrators who believed the NSF mandate was just a phase, who expected those in the disciplines to revolt and simply stop reviewing plans until the mandate went away, and who would not invest in research data management support because the “return on investment” was unclear. The latter position is fundamentally anti-intellectual and reveals a deep misunderstanding of the basic principles of research, which ideally begins with an unanswered question, not a financial statement. This is, unfortunately, perfectly consistent with the technocratic logic driving many university administrations.

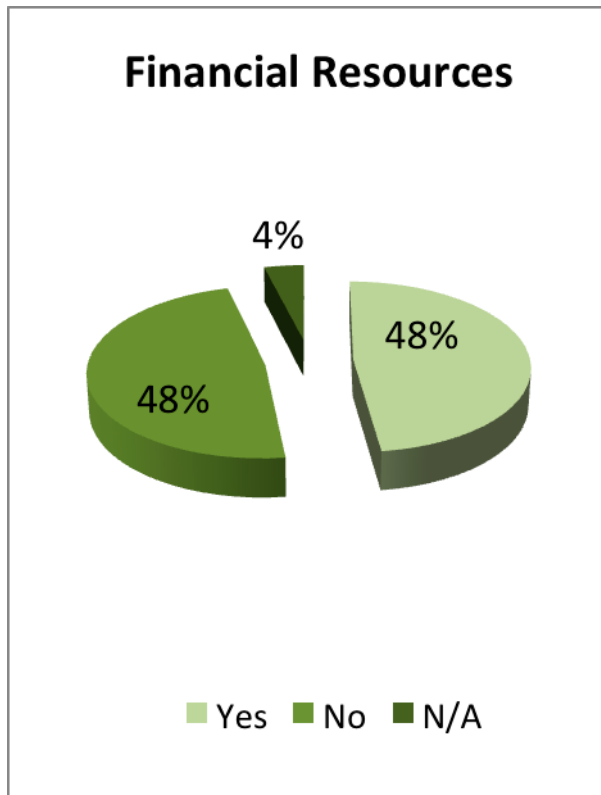


Fig. 9. Administrator responses to survey question, "Does your institution allocate financial resources for data management?"

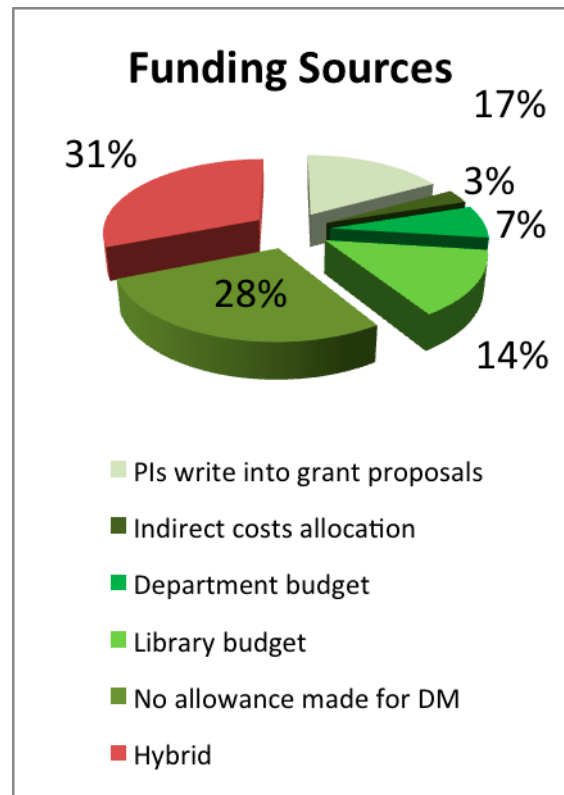


Fig. 10. Administrator responses to survey question, "From which sources are these funds drawn?" Multiple answer choices could be selected.

Conclusion: The Data Doldrums

At our final focus groups, conducted in June 2013 during the American Library Association annual conference in Chicago, we asked participants (mostly academic librarians) what message they most wanted us to take away from the discussion. Early in the research process, focus group participants had been anxious about the cost of implementing data management services, but eager to hear what was happening at other institutions and to share gossip and anecdotes about badly behaved principal investigators and administrators. In stark contrast, the atmosphere in the Chicago focus groups was noticeably subdued. Participants in these later focus groups most often used words like "worried," "anxious," and "stressed" to describe their feelings about data management services at their institutions.

One participant, a liaison librarian from a prestigious private research university, eloquently expressed her fear that library support for disciplines like philosophy, the humanities, and the soft social sciences would be left behind as university administrations and offices of research, library leadership, and funding agencies, including NEH-ODH, turn away from supporting traditional lines of scholarly inquiry in favor of data-driven (in particular, big data-driven) projects that are now "sexy." Without new funding to support research data management functions, the new focus on research data management will likely end up further overloading already overstressed

library budgets. This could potentially threaten to further weaken support for non-STEM (science, technology, engineering, and mathematics) research in favor of funding agency and university administration priorities that in many cases researchers in the STEM disciplines do not (yet) share.

How then do we finally understand the current status of research data management efforts in academia? In the two and a half years since NSF announced its data management plan requirement, academic libraries have scrambled to keep up with what continues to be perceived as another unfunded mandate. Returning to the nautical metaphors popular in discussions of big data, we are neither riding the wave nor being swamped by it. Rather, we may be *becalmed*, mired in the Sargassum of institutional inertia.

There was a significant degree of hope that the February 2013 memo of the Office of Science and Technology Policy nudging federal agencies to come up with a coherent strategy would spark some movement, but the August deadline for agency plans came and went with no public announcements. This silence was soon followed by the shutdown of the U.S. federal government in October 2013, an event that is all too emblematic of gridlock and being stuck in the doldrums. It now seems highly unlikely that vigorous and assertive prescriptions for research data management will be forthcoming from federal agencies, at least in the immediately foreseeable future.

In the absence of clear guidance from the federal agencies, university administrations are likely to fall back on the all too easy excuses for withholding resources from service providers—mainly libraries—believing that the requirement for a data management plan is a passing whim on the part of the agencies and that there is no point in investing time, money, and staff without a clear return on investment. Principal investigators, too, have room to doubt the seriousness of the data-sharing mandate and may simply continue to craft data management plans that reinforce the proprietary nature of their data rather than planning to make them available to be preserved, shared, and repurposed. And libraries may continue to try to meet the demands of both administrators and researchers with ever-shrinking financial resources—the equivalent of diligently polishing the decks and patching the sails in the vain hope that today the winds will return.

We should not allow our institutions to continue to drift in the data doldrums. To continue our metaphor, the promise of new lands is too great for us to accept remaining becalmed. But if we are to emerge from the doldrums, we will have to demonstrate stronger leadership and make greater efforts to work together within our institutions. Rather than waiting passively, we should take serious analytic notes from the small number of exemplar institutions in which librarians, researchers, and academic administrative leaders are collaboratively developing a shared agenda for research data management. Rowing out of the doldrums will require hard work, and we will have to row *together* to succeed. The question is not really

whether we will devote this effort to moving forward, it is rather *how long* we will collectively tolerate being becalmed. Our conclusion is that we should collectively *get moving*.

References

California Digital Library. 2013a. DMPTool. Available at <https://dmp.cdlib.org/>.

California Digital Library. 2013b. EZID. Available at <http://www.cdlib.org/services/uc3/ezid/index.html>.

California Digital Library. 2013c. University of California Curation Center. Available at <http://www.cdlib.org/services/uc3/index.html>.

Hermeneutica.ca. n.d. Voyant. Available at <http://voyant-tools.org/>.

Institute for Museum and Library Services. n.d. Specifications for Projects That Develop Digital Products. Available at <http://www.imls.gov/assets/1/AssetManager/DigitalProducts.pdf>.

Kelley, Michael. 2011. Librarians at University of Minnesota Make an Impact with Data Management Program. *Library Journal* (Aug. 9).

MIT Libraries. n.d.a. DSpace@MIT Available at <http://dspace.mit.edu/>.

MIT Libraries. n.d.b. *Self Help: Subject Guides: Data Management and Publishing*. Available at <http://libraries.mit.edu/guides/subjects/data-management/>.

Moen, William E., and Martin Halbert. 2012. *Support for Research Data Management among U.S. Academic Institutions: Results from a National Survey*. Denton: Texas Center for Digital Knowledge, College of Information, University of North Texas. Unpublished draft dated January 26, 2012; cited by permission. UNT Digital Library. <http://digital.library.unt.edu/ark:/67531/metadc181695/>.

Moretti, Franco. "Literary Lab Pamphlet 2: Network Theory, Plot Analysis." Stanford: Literary Lab, May 1, 2011. Available at <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf>.

National Endowment for the Humanities. Data Management Plans for NEH Office of Digital Humanities Proposals and Awards. Executive Summary. Available at http://www.neh.gov/files/grants/data_management_plans_2013.pdf

National Institutes of Health. 2003, February 26. Final NIH Statement on Sharing Research Data. NOT-OD-03-032. Available at <http://grants1.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.

National Science Foundation. n.d. Dissemination and Sharing of Research Data. Available at <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.

National Science Foundation. 2011. *Award and Administration Guide*. Chapter VI.D.4. Available at http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4.

National Science Foundation. 2013. *Grant Proposal Guide*, Chapter II.C.2.j. Available at http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp#dmp.

Pennsylvania State University Libraries. Digital Curation Services. Available at <http://www.libraries.psu.edu/psul/pubcur/curation.html>.

Purdue University Libraries. n.d. Data Curation Profiles Toolkit. Available at <http://datacurationprofiles.org/>.

Purdue University. 2013. Purdue University Research Repository. Available at <https://purr.purdue.edu/>.

Tufts University, 2013. Research Guides@Tufts. *Federal Funding Agencies: Data Management and Sharing Policies*. Available at <http://researchguides.library.tufts.edu/content.php?pid=167647&sid=1412586>.

University of California at San Diego. 2013a. RCI: Research Cyberinfrastructure. Available at <http://rci.ucsd.edu/>.

University of California at San Diego, The Library. 2013b. Research Data Curation Services. Available at <http://libraries.ucsd.edu/services/data-curation/index.html>.

University of Illinois at Urbana-Champaign. n.d. Illinois Digital Environment for Access to Learning and Scholarship. Available at <https://www.ideals.illinois.edu/>.

University of Illinois at Urbana-Champaign, University Library. Life Sciences Data Services, Data Management Plans. Available at <http://www.library.illinois.edu/lldata/dmp/dmp.html>.

University of Minnesota, University Libraries. 2011. Managing Your Data. Available at <http://www.lib.umn.edu/datamanagement>.

University of New Hampshire. (2012) Ownership and Management of Research Data. UNH.VII.C.1 <http://www.usnh.edu/olpm/UNH/VIII.Res/C.htm>.